The A-Not-B Error: Results from a Logistic Meta-Analysis

By: Stuart Marcovitch and Philip David Zelazo

Marcovitch, S., & Zelazo, P.D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, *70*, 1297-1313.

Made available courtesy of Wiley-Blackwell: http://www.wiley.com/

Abstract:
A meta-analysis of the A-not-B error was conducted using logistic regression on studies conducted before September 1997 (107 data points). An earlier meta-analysis by Wellman, Cross, and Bartsch revealed that age, delay between hiding and retrieval, and number of hiding locations were significant predictors of both the proportion of infants who searched correctly on B trials and the proportion of infants who searched perseveratively. The current analysis replicated these findings with two exceptions: (1) The number of trials at the A location was a significant predictor, and (2) the number of locations was a significant predictor of the proportion of infants who searched perseveratively, but *not* the proportion of infants who searched correctly. Implications of these findings are discussed and a quantitative version of a hierarchical competing-systems model of infant search is proposed.

Article:
INTRODUCTION
The standard A-not-B search task (Piaget, 1954) is a variant of the delayed-response paradigm introduced by Hunter (1913, 1917) to study the effects of delay on search behavior in human infants and nonhuman animals. In a delayed-response task, an object is conspicuously hidden in one of several locations according to a random schedule, and then, after a delay, the animal is allowed to search for it. In the A-not-B task, typically used with human infants, the number of hiding locations is usually restricted to two, and after a certain number of trials at one location (A), the object is hidden at the other location (B). Incorrect search at location A on B trials is termed an A-not-B, or perseverative, error.

Piaget used the A-not-B task to demonstrate the fragile nature of Stage IV infants' understanding of object permanence. However, contemporary researchers have recognized the utility of the paradigm as a measure of a wide range of cognitive and motoric abilities. Although some theorists have followed Piaget in proposing that the error is caused by conceptual difficulties such as a failure to understand specific properties of the desired object (e.g., Gratch, Appel, Evans, LeCompte, & Wright, 1974; Harris, 1983), others have emphasized basic cognitive processes such as deficits in short-term memory (e.g., Cummings & Bjork, 1983a, 1983b), the combination of memory deficits and lack of inhibitory control (e.g., Diamond, Cruttenden, & Niederman, 1994), or means-end behavior (e.g., Baillargeon, Graber, Devos, & Black, 1990; Diamond, 1991). In addition, some theorists have suggested that there is no need to implicate underlying cognitive processes at all, preferring instead to discuss perseveration in terms of the repetition of motoric schemes (e.g., Thelen & Smith, 1994). Thus, the A-not-B paradigm has provided a forum for research on infants' conceptual, mnemonic, inhibitory, and/or motoric abilities.

Since Piaget's seminal observations, a large number of studies have focused on manipulations that affect the likelihood of observing the A-not-B error. However, in several of these studies, the standard paradigm has been modified substantially. For example, to separate random from perseverative errors, researchers have used multiple hiding locations (e.g., Bjork & Cummings, 1984). This modification further necessitated changes to the hiding procedure that ensured that all hiding locations were covered simultaneously (see Diamond et al., 1994). Researchers have also varied the distance between hiding locations (e.g., Horobin & Acredolo, 1986) and the distinctiveness of hiding locations (e.g., Harris, 1974, Exp. 2). Additionally, multi-step retrieval procedures (e.g., Marcovitch, 1996; Zelazo, Reznick, & Spinazzola, 1998) have been used to elicit perseveration in children at older ages (e.g., as old as 28 months).

These differences in task presentation make it difficult to compare findings across studies without the use of statistical methods. Over a decade ago, Wellman, Cross, and Bartsch (1986) conducted a meta-analysis of results from 30 studies (standard multi-location analysis). The meta-analysis indicated that age, the length of delay between the hiding event and search, the number of hiding locations, and the distinctiveness of hiding locations all influence infants' search, whereas the number of A trials and the distance between hiding locations do not. Results regarding each of these variables constrain models of cognitive and behavioral development in infancy and early childhood.

For example, consider the finding that infants are more likely to make the A-not-B error as the delay increases. Longer delays should place greater demands on memory, so age-related changes in search may reflect improvements in short-term memory, both in terms of total capacity of the memory store and in terms of the duration of the memory trace. This conclusion is consistent with the results of individual studies. For example, Diamond (1985) has shown that as age increases, so does the length of delay required to commit perseverative errors on the A-not-B task.

The number of hiding locations and the distinctiveness of the hiding locations are also relevant to memory. Wellman et al.'s (1986) finding that an increase in the number of hiding locations increases the likelihood of correct search is difficult to explain in terms of memory models of A-not-B search (e.g., Goldman-Rakic, 1987). These models would presumably predict that an increase in the number of locations would increase the memory demand. On the other hand, the finding that the use of distinctive hiding locations leads to better search performance is consistent with the suggestion that the task assesses memory, because the use of distinctive hiding locations should attenuate memory demands.

Wellman et al.'s (1986) finding that the number of A trials does not predict infants' search behavior is surprising from the perspective of research on habit formation. In his Law of Exercise, Thorndike (1921) stated that repetition of a behavior under a set of conditions increases the probability that the behavior will be elicited under the same conditions in the future. Williams (1938) further demonstrated that the number of responses that rats required for extinction of a reinforced response increased as a function of the number of reinforcements. Finally, Hull theorized that "habit strength increases up to a maximal asymptote as the reinforcements increase" (Hull, Felsinger, Gladstone, & Yamaguchi, 1947, p. 239). In the context of the A-not-B task, each rewarded response to location A should increase the tendency to search at that location. On B trials, this tendency would be in conflict with representational information that encourages search at the B location (e.g., Diamond et al., 1994; Zelazo et al., 1998). If this were the case, then in the A-not-B task, infants may be unable to inhibit the tendency to search at location A. Age-related changes in search may reflect the growth of an inhibitory mechanism (e.g., Dempster, 1992; Diamond, 1991; Harnishfeger & Bjorklund, 1993). However, if performance on B trials is independent of the number of A trials, then there may be no need to implicate inhibitory processes (see Harris, 1986), and several approaches to infant search (e.g., Diamond et al., 1994; Zelazo et al., 1998) may need to be revised.

The theoretical importance of research on A-not-B search has sustained interest in the task in recent years (e.g., Hofstadter & Reznick, 1996; Smith, Thelen, Titzer, & McLin, 1999) and justifies further attempts to comprehend the myriad findings in this area. In this article, we present the results from a new meta-analysis that included studies conducted during the past decade. Moreover, we made use of techniques designed to improve on previous meta-analytic methodology. Specifically, the meta-analysis that we conducted was done using logistic regression in conjunction with cross-validation.

Wellman et al. (1986) used linear regression to analyze the data in their meta-analysis. This analysis assessed the influence of several variables on the proportion of children who made a perseverative error and the proportion of children who searched correctly on the first B trial. However, because proportional data are bounded (i.e., all data points must be between 0 and 1), the unbounded linear function may be inappropriate. For extreme values of an independent variable, the slope of the regression line changes to accommodate the data. On the other hand, logistic regression (Neter & Wasserman, 1974) offers an alternative to linear regression that is amenable to proportional data because the logistic function is bounded (i.e., the function is only defined between 0 and 1). The logistic function, L(*x*), is:

$$L(x) = 1/ [1 + \exp(—x)] \qquad (1)$$

The function's characteristic S-shaped curve allows it to accommodate extreme data points (see Discussion).

In the current analysis, the logistic regression was performed using a simple connectionist network with a back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). There are several advantages to conducting logistic regression analyses via a connectionist network rather than using a statistical package. The primary advantage lies in the ability to control for the complexity of the model and thus increase its generalizability, as will be discussed later. Most statistical packages use the conventional "sum of squares" to calculate the error term. However, for proportional data, it may be more appropriate to calculate the error term using less conventional functions (see Method). The back-propagation network has the advantage of allowing for the specification of *any* error term function. Finally, given that connectionist networks are increasingly common in the cognitive and developmental literature, they provide a convenient tool for regression analyses.

The network that we used consists of input nodes (the predictor variables), one output node (the computed value dependent variable), and the connections, or weights, between them. The network calculates the output by summing the input strengths (i.e., the multiplication of each input node with its respective weight). The output is the logistic of the calculated sum. The computed output is compared to the actual value of the dependent variable, and an error is established. The weights are initially random, but as the network encounters the set of data points, the weights change so as to minimize error. When error is at a minimum, the values of the weights are considered to be the regression coefficients (see Method for details). Note that if instead of using a logistic function we allowed the computed output to be only the sum of the input strengths, then this would be linear regression (i.e., the dependent variable would equal the sum of each predictor variable multiplied by its coefficient).

The primary objective in regression analysis is to determine the coefficients that best describe general trends in the data. The most appropriate regression curve is the one that accounts for the most variability in the data in the most parsimonious fashion. This curve should be able to predict results from novel data points more reliably than all other possible curves. This concept is known as generalizability.

A common problem with regression analysis is the lack of generalizability that results from overfitting the curve to the data; that is, finding coefficients that predict the relevant data set but do not generalize to other data. The strategy we used to test for generalizability was cross-validation. This involved holding back a subset of the data (i.e., the test set), conducting the regression on all data points except for those that were held back (i.e., on the training set), and then testing the generalizability of the regression curve on the test set (for further information on model validation, see Neter, Wasserman, & Kutner, 1990).

Another technique used to maximize generalizability in connectionist models is the minimization of complexity. Simply put, the more connections a network has, the more complex (and less generalizable) the network. A decay parameter can be introduced that will lower the weights proportionally to the inactivity of the connection. The result will be the nullification of infrequently used connections (see Method for details), which leads to a less complex and, therefore, more generalizable model.

The updated data set was subjected to a logistic meta-analysis. For the purpose of comparison, Wellman et al.'s (1986) "standard multilocation" data set was also analyzed, and both data sets were analyzed using both logistic and linear regression. The standard multilocation data set was chosen because the criteria for inclusion were most similar to the criteria chosen for determining inclusion in the current analysis. The results from these analyses were then used to motivate a comparison of alternative competing-systems models of infant search.

METHOD

Meta-analyses were conducted on two data sets. One data set, hereafter referred to as the original data set, was identical to the set used in Wellman et al.'s (1986) standard multilocation analysis. The other data set, hereafter referred to as the updated data set, was gathered according to slightly different criteria and included all relevant studies that were available when this article was prepared (September 1997). Both data sets were subjected to linear and logistic regression.

To increase the validity of the meta-analysis, it was necessary to establish criteria to determine the acceptability of particular studies. The criteria established by Wellman et al. (1986) for the standard multilocation analysis are presented in Table 1. These criteria were slightly modified to create the updated data set. The modifications were designed to eliminate search conditions in which methodological departures from the standard task might change the fundamental nature of the task. For example, we eliminated studies that used invisible hiding trials because in these cases infants had no information about the current location of the object and might respond randomly regardless of delay. See Table 2 for a list of the modifications to Wellman et al.'s criteria.

Once criteria were established, the next step was to compile the data for the current analysis. Several techniques were involved in the compilation of the updated data set. First, literature searches were conducted using keywords such as "A-not-B error," "object permanence," "perseveration," and "infant search." Second, programs from recent conferences (Society for Research in Child Development, 1995, 1997; International Conference on Infant Studies, 1996) were examined in an attempt to locate as-yet-unpublished studies involving the A-not-B paradigm. Finally, researchers who were known by us to have done recent work with the paradigm were contacted. In several cases, authors were contacted to disambiguate admissibility requirements. For studies prior to 1986, we compiled a list of data points similar to that of the original data set. Table 3 lists the 58 conditions used in the original data set, along with the 11 conditions that were deleted for the purpose of the updated data set. Table 4 lists the 60 additional conditions (including 15 conditions prior to 1986) that produced the 107 conditions in the updated data set.

Table 1 Admissibility Criteria for Studies in Wellman et al.'s (1986) Standard Multilocation Analysis

1. There must be at least two potential search locations.
2. The proportion of infants who erred on the first B trial must be reported.
3. The infant must be able to retrieve the toy on B trials.
4. Toys could not be changed systematically between the A trials and the B trials.
5. The object must not be visible while the infant searches.
6. Objects, not people, must be hidden.
7. Neither the infant nor the search array may be moved.
8. Hiding locations must be in the horizontal plane.
9. Distractions may not be present during hiding.
10. All covers and backgrounds must be identical.

Table 2 Modifications to Wellman et al.'s (1986) Standard Multilocation Admissibility Criteria

Modification	Justification
1. Covers do not have to be identical.	Distinctiveness of covers can be a predictor.[a]
2. Search during A trials must be active.	The effect of passive experience may differ from that of active experience.
3. Hiding must be conspicuous on both A and B trials.	Trials with invisible search are more likely to elicit random responding.
4. B trials must follow A trials directly.[b]	Extra delays (or intervening tasks) between A and B trials may weaken any effect of A trials.
5. There must be values for all variables tested.	Empty cells were not permitted in the regression analysis.
6. The minimum sample size was $n = 6$.	Arbitrary number chosen to balance the tradeoff between admissibility and reliability.

[a] Wellman et al. studied the effect of distinct versus identical covers, but not in the standard multilocation analysis.
[b] The delay between A and B trials should be no longer than the delay between two A trials and there should be no intervening tasks between A and B trials.

The six predictor variables tested in the current analysis of the updated data set were: age, number of hiding locations, distance between locations, delay between hiding and search, number of A trials, and whether or not the covers or backgrounds were distinct. Table 5 gives the mean, standard deviation, and ranges for each variable in the current analysis. Where applicable, these statistics are also listed for the variables used in the analysis of the original data set. As a control, a dummy variable that consisted of random numbers was assigned to each data point. Control variables are necessary to ascertain that results from the meta-analysis are not due to chance outcomes. In the analysis of the original data set, only the variables tested by Wellman et al. (1986) were tested (i.e., age, delay, number of locations, and number of A trials).

Following Wellman et al., the dependent measures for both data sets were the proportion of infants searching at B (correct) on the first B trial, and the proportion searching at A (perseverative) on the first B trial. There is a logical dependency between these two variables because a perseverative error is necessarily incorrect. It is not an absolute dependency, however, because incorrect search is not necessarily perseverative when more than two hiding locations are used. Nonetheless, there would be some redundancy if both measures were reported separately without correcting for the overlap. To address this issue, a first analysis was done using the proportion of infants searching correctly as the dependent variable. Then, a second analysis was done using the proportion of infants searching perseveratively, but for these analyses, the proportion of infants searching correctly was used as a predictor variable. Thus, any variables found to be significant predictors of the proportion of infants who searched perseveratively would be variables that predict perseverative search above and beyond correct search.

The analyses were conducted using a neural network with a back propagation learning algorithm (Rumelhart et al., 1986). The input strength (x) was the sum of each input value multiplied by its respective weight, that is $x = \Sigma_j (\text{input}_j) \text{ X } w_j$. The computed output ($y$) was: $y = x$ for the linear analysis, and $y = L(x)$ (see Equation 1) for the logistic analysis. The input units corresponded to the independent variables and a bias unit (i.e., a unit with value fixed at 1). All input values were normalized; that is, the *z* score rather than the raw score was used as the input. To obtain a measure of error, the value of the computed output was compared with the value of the dependent variable. Prior to each analysis, all weights were randomized. Furthermore, no hidden units were used in any of the analyses.[1]

Linear regression. The linear analysis was intended to mimic the meta-analysis conducted by Wellman et al. (1986). Thus, techniques that might be used to improve generalizability of the model, such as cross-validation and minimizing complexity cost, were not used. For the linear analysis, the error function to be minimized was: $E = \Sigma 1/2(d - y)^2$, where *y* is the computed output and *d* is the actual proportion (that is, proportion of infants searching correctly on the first B trial or proportion of infants searching perseveratively on the first B trial). The final weights were the linear coefficients for the normalized data.

If the value of a coefficient was zero, we assumed that the variable did not play a role in predicting the dependent variable. To determine whether the coefficients were significantly different from zero, a boot-strapping method was employed. That is, for each data set in each analysis, 20 additional sets were created by random selection with replacement. These sets contained as many data points as the original set. For example, in the original data set, which consisted of 58 data points, all 20 additional sets also contained 58 points. By conducting the analysis on each of the additional data sets, 20 estimates for each weight were generated. After all 20 additional sets were analyzed, the sample mean and the sample standard deviation of the weight estimates were calculated. If zero did not lie within two sample standard deviations of the sample mean, then the value of the weight was judged to be significantly greater than (or less than) zero and the variable was considered to be a significant predictor.[2] The analysis was then conducted on the original data set using only the significant predictor variables. The use of this method guards against the possibility of the network converging at a local, rather than the absolute, minimum. Multiple simulations at different initial random weights diminishes the effect, if any, of local minima.

Logistic regression. The logistic analysis was intended to be a statistical improvement on previous analyses. In that vein, other methodological improvements designed to promote the generalizability of the model were also implemented, including cross-validation and minimizing complexity cost.

The logistic network was identical to the linear one, with two exceptions. First, a logistic instead of a linear function was used to compute the output. Second, a different error function was minimized, namely:

$$E = d \times 1n(1/y) + (1 - d) \times 1n[1/(1 - y)] \qquad (2)$$

Recall that *d* is the actual proportion (proportion of infants searching correctly or proportion of infants searching perseveratively), while *y* is the computed output. This error function, which is only defined when $y > 0$, interprets the outputs as probabilistic decisions, and thus is particularly amenable to proportional data (see van Camp, 1996). Note that while this error function is preferred for proportional data, it cannot be used with linear regression because the computed output for the linear regression is not always greater than zero.

The first step involved in our logistic analysis was to hold back data for the purpose of cross-validation. Eight subsets of the original data set and 10 subsets of the updated data set were created with the following restrictions: Each subset contained approximately equal numbers of data points, with every data point randomly assigned to one and only one subset. For each subset, the model was trained on the complete data set minus the subset (i.e., on the training set), and subsequently tested on the subset (i.e., on the test set). The error term that resulted from the test set provided an index of generalizability for the model (i.e., the lower the error term, the more generalizable the model).

Table 3 Study Conditions Used in Wellman et al.'s (1986) Meta-Analysis

Study	Condition
1. Appel & Gratch (1984)	9 months, toy-toy condition
2. Appel & Gratch (1984)	12 months, toy-toy condition
3. Bjork & Cummings (1984, Exp. 1)	5 locations
4. Bjork & Cummings (1984, Exp. 2)	2 locations
5. Bjork & Cummings (1984, Exp. 2)	5 locations
6. Butterworth (1977)	8 months, 3 A trials, traditional
7. Butterworth (1977)	8 months, 5 A trials, traditional
8. Butterworth (1977)	9 months, 3 A trials, traditional
9. Butterworth (1977)	9 months, 5 A trials, traditional
10. Butterworth (1977)	10 months, 3 A trials, traditional
11. Butterworth (1977)	10 months, 5 A trials, traditional
12. Butterworth, Jarret, & Hicks (1982, Exp. 1)	Condition III, 8 months
13. Butterworth, Jarret, & Hicks (1982, Exp. 1)	Condition III, 9 months
14. Butterworth, Jarret, & Hicks (1982, Exp. 1)	Condition III, 10 months
15. Cummings & Bjork (1983a)	5 locations, end-middle
16. Cummings & Bjork (1983a)	5 locations, end-end
17. Cummings & Bjork (1983a)	5 locations, middle-end
18. Cummings & Bjork (1983b)	3 locations
19. Cummings & Bjork (1983b)	5 locations
20. Cummings & Bjork (1983b)	6 locations
21. Evans (1973)[a]	2 A trials, passive experience
22. Evans (1973)	2 A trials, active experience
23. Evans (1973)[a]	5 A trials, passive experience
24. Evans (1973)	5 A trials, active experience
25. Evans & Gratch (1972)	Same toy
26. Frye (1980, Exp. I)[b]	Control
27. Gratch, Appel, Evans, LeCompte, & Wright (1974)	0 s delay
28. Gratch, Appel, Evans, LeCompte, & Wright (1974)	1 s delay
29. Gratch, Appel, Evans, LeCompte, & Wright (1974)	3 s delay
30. Gratch, Appel, Evans, LeCompte, & Wright (1974)	7 s delay
31. Gratch, Appel, Evans, LeCompte, & Wright (1974, p. 74)	0 s delay
32. Harris (1973, Exp. I)[c]	Change location, same response
33. Harris (1973, Exp. II)[d]	0 s delay
34. Harris (1973, Exp. II)[d]	5 s delay
35. Horobin & Acredolo (1986)	2 locations, narrow distance
36. Horobin & Acredolo (1986)	2 locations, wide distance
37. Horobin & Acredolo (1986)	6 locations
38. Landers (1971)	Little active experience
39. Landers (1971)	Much active experience
40. Landers (1971)[a]	Passive experience
41. Schuberth, Werner, & Lipsitt (1978)	Same toy
42. Sophian (1985)	2 locations, traditional
43. Sophian (1985)	3 locations, traditional
44. Sophian & Sage (1985)	9 months, identical locations
45. Sophian & Sage (1985)	16 months, identical locations
46. Sophian & Wellman (1983, Exp. I)[a]	9 months, visible hiding
47. Sophian & Wellman (1983, Exp. I)[e]	9 months, finding at A
48. Sophian & Wellman (1983, Exp. I)	9 months, 1 trial at A
49. Sophian & Wellman (1983, Exp. I)	9 months, 3 trials at A
50. Sophian & Wellman (1983, Exp. I)[a]	16 months, visible hiding
51. Sophian & Wellman (1983, Exp. I)[e]	16 months, finding at A
52. Sophian & Wellman (1983, Exp. I)	16 months, 1 trial at A
53. Sophian & Wellman (1983, Exp. I)	16 months, 3 trials at A
54. Sophian & Yengo (1985)	Hidden object: 2 locations
55. Sophian & Yengo (1985)	Hidden object: 3 locations
56. Webb, Massar, & Nadolny (1972, 16 months)	Cups
57. Webb, Massar, & Nadolny (1972, 16 months)	Toy chest
58. Webb, Massar, & Nadolny (1972, 16 months)	Quart container

Note: Several conditions were deleted from the analysis of the updated set because: [a] search during A trials was not active, [b] B trials did not directly follow A trials, [c] the search array was moved, [d] proportion of infants who erred on the first B trial was not reported, [e] hiding was not conspicuous on all trials. Adapted from "Infant Search and Object Permanence: A Meta-Analysis of the A-Not-B Error," by H. M. Wellman, D. Cross, and K. Bartsch, 1986, *Monographs of the Society for Research in Child Development*, *51*, pp. 8–9.

Table 4 Additions to the Original Data Set

Study	Condition
1. Bell & Adams (1997)	Reaching condition
2. Bigelow, MacDonald, & MacDonald (1995)	Small object, 8 months
3. Bigelow, MacDonald, & MacDonald (1995)	Small object, 10 months
4. Bigelow, MacDonald, & MacDonald (1995)	Small object, 12 months
5. Bigelow, MacDonald, & MacDonald (1995)	Large object, 8 months
6. Bigelow, MacDonald, & MacDonald (1995)	Large object, 10 months
7. Bigelow, MacDonald, & MacDonald (1995)	Large object, 12 months
8. Bremner (1978)[a]	Traditional, distinctive covers
9. Bremner & Bryant (1977)[a]	Traditional, distinctive backgrounds
10. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition I, 8 months, distinct backgrounds and covers
11. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition I, 9 months, distinct backgrounds and covers
12. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition I, 10 months, distinct backgrounds and covers
13. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition II, 8 months, distinct backgrounds
14. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition II, 9 months, distinct backgrounds
15. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition II, 10 months, distinct backgrounds
16. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition III, 8 months, distinct covers
17. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition III, 9 months, distinct covers
18. Butterworth, Jarret, & Hicks (1982, Exp. 1)[a]	Condition III, 10 months, distinct covers
19. Diamond, Cruttenden, & Neiderman (1994)	Slits, simultaneous covering
20. Diamond, Cruttenden, & Neiderman (1994)	Slits, attention drawn to correct well
21. Diamond, Cruttenden, & Neiderman (1994)	Covers, attention drawn to correct well
22. Diamond & Goldman-Rakic (1989)	Human infants, 7 months, 2 s delay
23. Diamond & Goldman-Rakic (1989)	Human infants, 7.5 months, 2 s delay
24. Diamond & Goldman-Rakic (1989)	Human infants, 8 months, 2 s delay
25. Diamond & Goldman-Rakic (1989)	Human infants, 8.5 months, 2 s delay
26. Diamond & Goldman-Rakic (1989)	Human infants, 9 months, 2 s delay
27. Diamond & Goldman-Rakic (1989)	Human infants, 9.5 months, 2 s delay
28. Diamond & Goldman-Rakic (1989)	Human infants, 10 months, 2 s delay
29. Diamond & Goldman-Rakic (1989)	Human infants, 10.5 months, 2 s delay
30. Diamond & Goldman-Rakic (1989)	Human infants, 11 months, 2 s delay
31. Diamond & Goldman-Rakic (1989)	Human infants, 11.5 months, 2 s delay
32. Diamond & Goldman-Rakic (1989)	Human infants, 12 months, 2 s delay
33. Diamond & Goldman-Rakic (1989)	Human infants, 7.5 months, 5 s delay
34. Diamond & Goldman-Rakic (1989)	Human infants, 8 months, 5 s delay
35. Diamond & Goldman-Rakic (1989)	Human infants, 11 months, 5 s delay
36. Diamond & Goldman-Rakic (1989)	Human infants, 12 months, 5 s delay
37. Diamond & Goldman-Rakic (1989)	Human infants, 12 months, 10 s delay
38. Harris (1973, Exp. III)[b]	Condition "1"
39. Harris (1973, Exp. III)[b]	Condition "2"
40. Hofstadter & Reznick (1996, Exp. 1)	Reach condition, 7 months
41. Hofstadter & Reznick (1996, Exp. 1)	Reach condition, 9 months
42. Hofstadter & Reznick (1996, Exp. 1)	Reach condition, 11 months
43. Mistry, Colombo, & Saxon (1997)	"No Cue" condition
44. Roberts (1997)	8 months
45. Roberts (1997)	10 months
46. Smith, Thelen, Titzer, & McLin (1997, Exp. 1)	Standard, 8 months
47. Smith, Thelen, Titzer, & McLin (1997, Exp. 1)	Standard, 10 months
48. Smith, Thelen, Titzer, & McLin (1997, Exp. 1)	Train at C, 10 months
49. Smith, Thelen, Titzer, & McLin (1997, Exp. 4)	Tap A, 10 months
50. Smith, Thelen, Titzer, & McLin (1997, Exp. 5)	Tap A, 10 months
51. Smith, Thelen, Titzer, & McLin (1997, Exp. 5)	Tap B, 10 months
52. Smith, Thelen, Titzer, & McLin (1997, Exp. 6)	Visual distraction, 10 months
53. Sophian & Sage (1985)[a]	9 months, distinct locations
54. Sophian & Sage (1985)[a]	16 months, distinct locations
55. Wishart (1987)	Nonretarded group, task 1, 45 months
56. Wishart (1987)	Nonretarded group, task 1, 45.5 months
57. Wishart (1987)	Nonretarded group, task 1, 46 months
58. Wishart (1987)	Nonretarded group, task 1, 46.5 months
59. Wishart (1987)	Nonretarded group, task 1, 47 months
60. Wishart (1987)	Nonretarded group, task 1, 47.5 months

[a] These analyses prior to 1986 were not included in the Wellman et al. analysis because covers and/or backgrounds were not identical.
[b] It is unclear why these conditions were not used.

Table 5 Means, Standard Deviations, and Ranges for the Variables in Each Data Set

	Original			Updated		
Variable	*M*	*SD*	Range	*M*	*SD*	Range
Age (months)	10.08	2.46	8–16	11.82	8.58	7–47.5
Number of hiding locations	2.72	1.12	2–6	2.52	1.19	2–7
Distance between locations (in centimeters)				20.87	8.34	4–48
Delay (s)	3.02	2.23	0–15	2.85	2.14	0–15
Number of A trials	3.03	1.80	1–9	2.46	1.34	1–9
Distinctiveness of covers and/or backgrounds[a]				.12	.33	0–1

[a] Identical covers and backgrounds were assigned a value of 0, while distinct covers and/or backgrounds were assigned a value of 1.

Cross-validation in conjunction with minimizing complexity cost can produce a model that maximizes generality. Minimizing complexity cost is achieved by introducing a "weight decay" parameter (λ) into the model. The larger the value of λ, the more likely it will nullify small weights, allowing the model to have fewer connections (i.e., to be less specialized) and therefore to be more generalizable. By changing the parameters, it is possible to find the value of λ that will lead to the most generalizable model, as determined by the error term of the test set. For each training set, different values of λ were introduced. The error terms for all test sets were summed to give a total error term for the data. The value of λ at which the total error term was at the minimum was deemed to produce the most generalizable model.

Once the value of λ was established, the bootstrapping method described earlier was used to determine which variables were significant predictors of the dependent variable. After the significant predictors were discovered, the network was rerun with the complete data set, the appropriate λ, and only the significant independent variables as inputs.

RESULTS

Original data set. Our first goal was to replicate the findings from Wellman et al.'s (1986) meta-analysis. To do so, we analyzed the standard multilocation data set (see Table 3), and tested the same predictors as did Wellman et al.: age, delay, number of locations, and number of A trials. However, instead of using year of publication as a control variable, we assigned random numbers to each data point. Wellman et al. (1986) found that age, delay, and number of locations were significant predictors whether the dependent variable was the proportion of infants who perseverated or the proportion of infants who were correct on the first B trial. The regression lines from the 1986 analysis were:

perseveration = 10.79 - .24[age] + .26[delay]
- 2.01[locations] (p. 23) (3)

correct = 8.77 + .24[age] - .30[delay]
+ .47[locations] (p. 25) (4)

The linear analysis that we conducted on the original data set, using proportion of infants who searched correctly as the dependent variable, confirmed Wellman et al.'s (1986) results; that is, age, delay, and number of locations were all significant predictors. An increase in age or the number of locations increased the proportion of infants who were correct on the first B trial, while longer delays decreased it. However, when the dependent measure was the proportion of infants who searched perseveratively at B, age, number of locations, and correct behavior were found to be significant predictors. Thus, age and number of locations predicted perseverative behavior above and beyond what they predict for correct search. The regression lines from our analysis were:

$$\text{correct} = .07 + .05[\text{age}] - .06[\text{delay}] + .04[\text{locations}] \quad (5)$$

$$\text{perseveration} = 1.15 + .01[\text{age}] - .12[\text{locations}] - 1.05[\text{correct}] \quad (6)$$

Note that the regression line for correct behavior (Equation 5) was not an exact replication of the line discovered by Wellman et al.

The logistic analysis on the original data set was consistent with our linear analysis. Age, delay, and location were found to be significant predictors when the dependent variable was the proportion of infants who searched correctly. When the dependent variable was the proportion of infants who searched perseveratively, age, location, and correct search were significant predictors. The regression functions for our logistic regression were:

$$\text{correct} = L(-1.84 + .24[\text{age}] - .30[\text{delay}] + .13[\text{locations}]),\ \lambda = .7 \quad (7)$$

$$\text{perseveration} = L(3.35 + .09[\text{age}] - .74[\text{locations}] - 5.36[\text{correct}]),\ \lambda = 0 \quad (8)$$

Updated data set. The linear analyses of the updated data set yielded age and delay as significant predictors of the proportion of infants who were correct on the first B trial, but it also yielded number of A trials and failed to yield number of locations. However, for perseverative performance, only number of locations and delay were found to be significant predictors above and beyond predictions of correct performance. The regression lines were:

$$\text{correct} = .65 + .01[\text{age}] - .05[\text{delay}] - .04[\text{A trials}] \quad (9)$$

$$\text{perseveration} = 1.18 + .01[\text{delay}] - .11[\text{locations}] - .96[\text{correct}] \quad (10)$$

Logistic analysis on the updated data set extended the general conclusions of the linear analysis. When the dependent variable was the proportion of infants who searched correctly, distance between locations was found to be a significant predictor (in addition to age, delay, and number of A trials). For the proportion of infants who searched perseveratively, age, delay, number of locations, and distance between locations were found to predict perseveration above and beyond correct behavior. The regression lines were:

$$\text{correct} = L(-1.25 + .21[\text{age}] - .25[\text{delay}] - .17[\text{A trials}] + .02[\text{distance}]),\ \lambda = .1 \quad (11)$$

$$\text{perseveration} = L(4.50 - .02[\text{age}] + .05[\text{delay}] - .76[\text{locations}] - .01[\text{distance}] - 5.26[\text{correct}]),\ \lambda = 0 \quad (12)$$

DISCUSSION

In their meta-analysis of the A-not-B error, Wellman et al. (1986) found that only age, delay, and number of locations were reliable predictors of infants' performance on the A-not-B search task. Our aim was twofold: to update the Wellman et al. data set by adding recent studies, and to introduce methodological improvements to the meta-analytic technique. The latter aim was achieved in three ways. First, additional admissibility criteria were established so as to include a more homogeneous set of studies. Second, we used a logistic function, which

is more suitable for proportional data. Third, data were held back in order to test the generalizability of the model.
The superiority of logistic regression for the analysis of proportional data can be seen in Figure 1, where the simple case of univariate regression is illustrated. Recall that the linear function is unbounded. The upper left panel of Figure 1 depicts the proportion of infants who search correctly on the first B trial as a function of age (measured as a *z* score). The addition of extreme data points, specifically the ages reported by Wishart (1987), which ranged from 45 to 47.5 months, changes the slope of the regression line drastically as shown in the upper right panel of Figure 1. However, the bounded logistic function, with its characteristic S-shaped curve, accommodates extreme data points, as can be seen by the lower panels of Figure 1.

Although in general the linear and logistic regressions identified the same predictors, logistic regression appeared to be more sensitive. For the updated data set, additional variables were found to be predictors of both correct and perseverative behavior (compare Equations 9 and 10 with Equations 11 and 12). Also, the theoretical advantage of logistic regression is evident in the quantitative differences in the results. For example, when the dependent measure was the proportion of infants who searched correctly (Equations 9 and 11), the coefficient for age is relatively small for linear regression (.01) versus logistic regression (.21). This is probably due to the inclusion of extreme age scores in the updated data set (e.g., those from Wishart, 1987). The logistic analysis provides a more appropriate fit to the data and thus yields a more accurate estimate of the relative strength of each predictor variable.

The predictor variables did vary depending on whether correct search or perseveration was used as the dependent measure. In the first set of analyses, which examined the proportion of infants who searched correctly, interpretation of these regression functions is straightforward: The presence of a given variable indicates that its effect is nonzero, and the coefficient of the variable describes both the direction of the effect and its relative strength. In the second set of analyses, however, interpretation of the regression functions can be complicated. In these analyses, the proportion of infants who searched correctly was used to predict the proportion of infants who searched perseveratively, in an attempt to discover the unique influence of the independent variables on perseverative behavior above and beyond correct behavior. Several things should be noted. First, the absence of a variable does not necessarily imply that the variable does not predict perseverative behavior; rather it implies that it does not predict perseveration over and above any predictions of correct behavior. So, if a variable predicts correct behavior (e.g., Equation 9, A trials) but not perseverative behavior (e.g., Equation 10, no A trials), this implies that the variable trials predicts perseverative behavior to the same degree that it predicts correct behavior.[3] Second, the presence of a variable implies prediction of perseverative behavior to a degree different than the prediction of correct behavior. This difference can be qualitative and/or quantitative. If the variable is not a significant predictor of correct behavior but *is* a significant predictor of perseverative behavior (e.g., number of locations in the updated data set, see Equations 11 and 12), then this difference can be called qualitative. This implies that the variable predicts perseverative, but not correct, behavior. However, if the variable is a significant predictor of both correct and perseverative behavior, the difference can be seen as quantitative (i.e., the variable's prediction of correct behavior does not entirely capture the variable's prediction of perseverative behavior).

Quantitative differences could arise for two reasons. First, the reason may be psychological in nature. For example, consider the effect of age. In the updated data set using logistic regression, age was found to predict correct behavior (+.21, Equation 11) to a different degree than perseverative behavior (—.02 beyond the prediction of correct behavior, Equation 12). One could argue that this difference reflects a meaningful developmental process. As infants age, they make more correct choices, and thus fewer overall errors (Equation 11). The age-related decrease in the likelihood of making perseverative errors, as opposed to nonperseverative errors (Equation 12), may indicate that when older infants do err, their errors are more likely (relative to younger infants) to reflect inattention than lack of inhibition. In other words, this quantitative difference may reflect a developmental change in the basis of infants' errors and thus might usefully be explored in future research.

The other reason a quantitative difference may arise is statistical in nature. Although the two dependent measures were logically dependent, they were not perfectly dependent, which would likely lead to slightly different regression coefficients due to chance alone. That is, even with no further psychological explanation, it is possible to find variations in predictive strength due to chance. In these situations, however, the differences between the coefficients should be small. In summary, a quantitative difference may reflect meaningful developmental processes, random variation, or both. To disambiguate the cause for any difference, one must consider the magnitude of the difference.

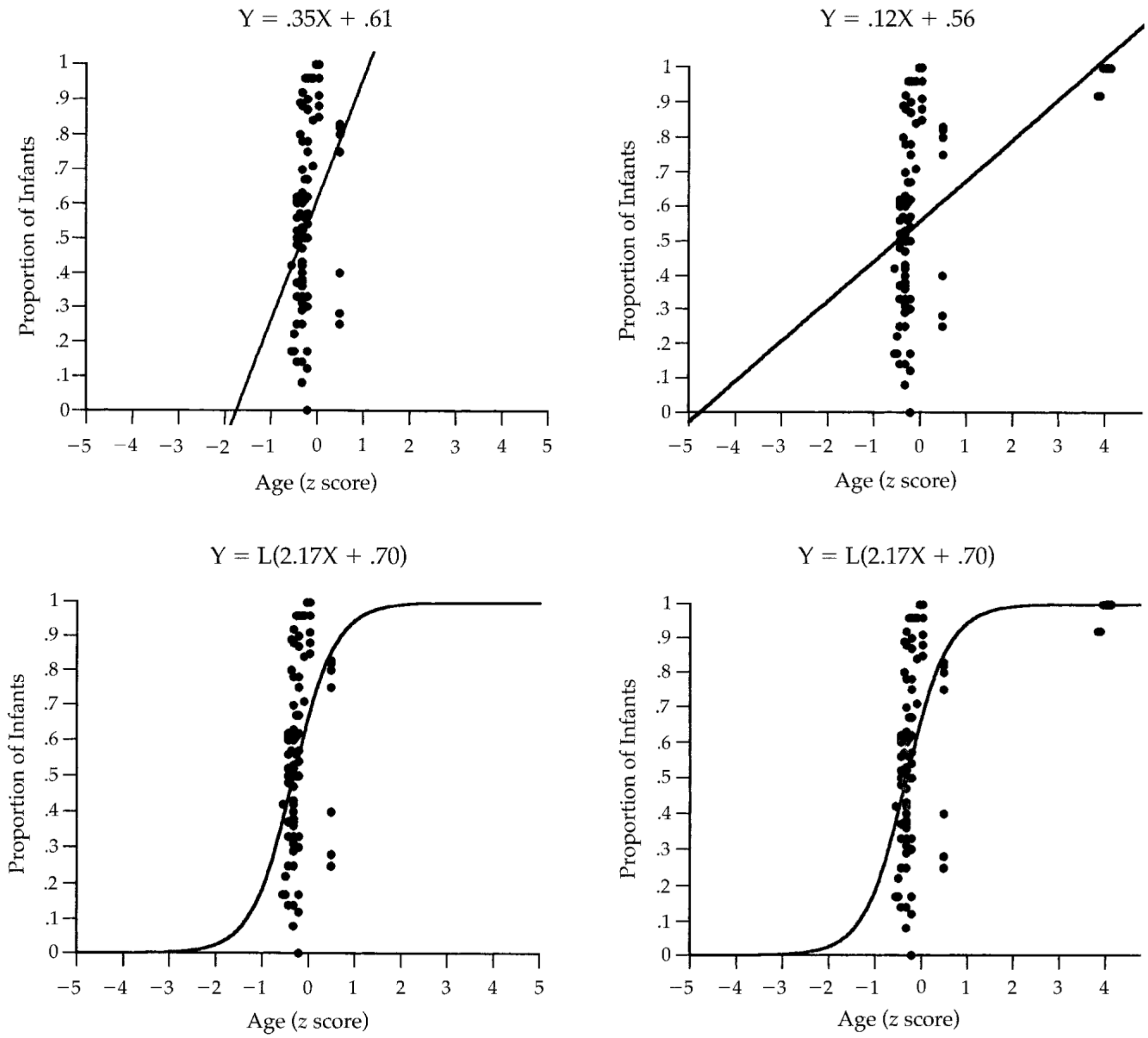

Figure 1 **Best linear (upper panels) and logistic (lower panels) regression curves when age is the independent variable and pro portion of infants correct is the dependent variable. Note that the linear regression curve changes to accommodate extreme dat: points (upper right panel), while the logistic regression curve does not (lower right panel).**

Original data set. Both linear and logistic regression on the original data set revealed that age, delay, and number of locations were significant predictors of infants' search when the dependent measure was the proportion of infants who searched correctly. However, age and number of locations predicted the proportion of infants who searched perseveratively beyond what these variables predicted about correct search behavior. For both variables, the difference between regression coefficients is quantitative.

Linear regression on the original data set (correct) was intended to be a replication of the Wellman et al. (1986) meta-analysis, and thus we would expect similar, if not identical, regression lines. It is clear by comparing Equation 3 to Equation 5 that there are quantitative differences between the regression lines (although the significant predictor variables were the same). The differences between the regression lines may be due to Wellman et al.'s arc sine transformation of the dependent variable (see Wellman et al., pp. 12–13). However, the regression lines that they report (Equations 3 and 4) cannot be interpreted directly, in part because the authors do not provide information regarding the units that were used for each variable, and do not indicate whether the data were re-transformed. For example, if the plausible values of age = 9 (months), delay = 3 (s), and locations = 2, are substituted for the variables in Equation 4, then the predicted proportion of infants who search correctly is 10.97, which is an impossible result. Even if the dependent variable is assumed to be the arc sine of the proportion of infants who search correctly— that is—arc sine(correct) = 10.97—then the predicted proportion of infants who search correctly is -1.00, which is also an impossible result. Due to the ambiguities concerning the Wellman et al. regression lines, we are reluctant to speculate on the nature of the differences between their findings and the results of our re-analysis of the original data set.

Updated data set. The linear analysis on the updated data set revealed that age, delay, and number of A trials were significant predictors of the proportion of infants who responded correctly on the first B trial. In addition to these variables, number of locations and delay were also found to be significant predictors when the dependent variable was the proportion of infants who responded perseveratively on the first B trial. Although the effect of delay was a relatively small, quantitative difference, the effect of number of hiding locations was a larger, *qualitative* difference. This implies that the number of hiding locations predicts perseverative, but not correct behavior. The logistic analysis on the updated set yielded similar results, with a few exceptions. In this analysis, distance was also a predictor of both correct search behavior and perseverative behavior.

The finding that age and delay were significant predictors of infants' search behavior on the first B trial was consistent with Wellman et al.'s (1986) conclusions. The results from studies that were specifically designed to assess the influence of these factors confirm these findings. For example, Sophian and Wellman (1983, Exp. 1) tested 9- and 16-month-old infants and found that with all other factors held constant, as age increased so did the proportion of infants who searched correctly on the first B trial. Similarly, Gratch et al. (1974) varied delay between hiding and search and found that as delay increased, the proportion of infants who searched correctly decreased.

On the basis of their analyses, and confirmed by our analyses using their data set, Wellman et al. (1986) concluded that number of A trials did not predict the outcome on the first B trial. Researchers, however, have typically used a small number of A trials. Thus, nearly all of the data points have three or fewer A trials (41 of 58 for the Wellman et al. data set, 90 of 107 for the updated data set). Because of the low number of A trials in studies used in the sample, it is possible that 58 data points were not sufficient to reveal an effect of number of A trials.

Direct tests of the effect of number of A trials have been promising. Landers (1971) presented infants with either low active experience (two A trials) or high active experience (8 or 10 A trials). Two measures were reported: the proportion of infants who perseverated, and the error run, which is the number of B trials needed until a criterion of a single correct response was met. The proportion of infants who perseverated on the first B trial was .71 for low active experience, and .86 for high active experience. Furthermore, 9-monthold infants given two pre-switch trials had a median error run of 1.5 trials (range = 0–8 trials) and those given eight or ten pre-switch trials had a median error run of 6.5 trials (range = 0–21 trials). Both measures suggest that there may be an effect of A trials; however, the error run may be more sensitive to any effects because it is a quantitative measure of habit as opposed to a categorical one.

Wellman et al. (1986) also concluded that as the number of locations increased, the proportion of infants who searched perseveratively decreased, while the proportion of infants who searched correctly increased. This implies that with three or more hiding places, infants perseverate less because they search correctly at B more often. In our analyses on the updated data set, however, both linear and logistic regression revealed that as the number of locations increases, the proportion of infants who search perseveratively decreases, but the proportion of infants who search correctly does not change. This finding is also supported by the magnitude of the quantitative differences for the location variable in the original data set (.12 for linear, .74 for logistic). The magnitudes of these coefficients are much larger than the other quantitative differences reported. Thus, we can be reasonably confident that this difference is due to underlying psychological processes as opposed to random variation. These results imply that with three or more hiding places, infants perseverate less because they search at alternative locations (neither A nor B) more often.

Several explanations may account for this finding. One possibility is the "null account," which states that perseverative errors are merely an artifact of the two-location paradigm (see Bjork & Cummings, 1984). On any A-not-B task, infants will either search correctly or incorrectly. If there are more than two locations, the incorrect responses will be randomly distributed among all the incorrect locations (including A). As a result, the proportion of infants who perseverate diminishes with the increase in number of locations; however, the proportion of infants who are correct will not change.

Another possibility is the "competing-systems" account, which is broadly consistent with a number of models in the literature (cf. Diamond, 1991; Luria, 1961; Munakata, 1997; Thelen & Smith, 1994; Zelazo & Frye, 1997; Zelazo et al., 1998). One version, outlined by Marcovitch (1996) and Zelazo et al. (1998), assumes that two hierarchically organized dissociable systems are involved in search tasks: a response-based system and a conscious representational system that has the potential to control the response-based system. The response-based system is activated by motor experience and the influence of the system is directly proportional to the number of reaches made to a particular location, consistent with findings from research on association learning (Hull, 1943) and skill acquisition (Anderson, 1993). On the other hand, the representational system is activated by the infant's conscious representation of the location of the hidden object (see Zelazo & Zelazo, 1998, for further discussion). The relative influences of the two systems determine the infant's behavior and, with age, the influence of the representational system increases. On the A trials, the two systems work in concert (i.e., both systems guide the infant toward location A). On the first B trial, however, the two systems will be in conflict and the response-based system guides the infant toward A while the representational system guides the infant toward B. A perseverative error is more likely to occur when the influence of the response-based system causes the infant to search erroneously toward the A location. If there were only two locations, all errors would be at location A. With three or more locations, however, the errors would be distributed among the locations that are between locations A and B. Thus, as the number of locations increases, the percentage of errors that actually result in search at A decreases, but correct search at B is unaffected.

To test whether the null account or the competing-systems account provides a better explanation, one would need to analyze the results from research that uses a multilocation A-not-B task and reports the distribution of searches at each location on the first B trial. If the distribution of searches at each of the incorrect locations occurred with equal frequency, the result would support the null account. However, if the frequency of searches toward the A location were higher than the frequency of searches away from the A location (i.e., on the other side of location B), this finding would provide evidence for a competing-systems account.

Diamond et al. (1994) conducted an A-not-B search experiment with seven hiding locations and reported the distribution of searches at each location. In their study, the A location was location #2, and the B location was location #5. Figure 2 shows the distribution of searches at each location. Incorrect searches were more likely to occur toward the A location (locations #3 and #4, 36% of searches) than away from the A location (locations #6 and #7, 15% of searches). This evidence supports a competing-systems account. Diamond et al. explained their results in terms of one such account, a memory plus inhibition account. The memory of the object's location directs search to location B. The inability to inhibit an action that led previously to a rewarded response allows for habit to direct search to location A. Memory plus inhibition (plus habit) jointly determine where the infant will search.

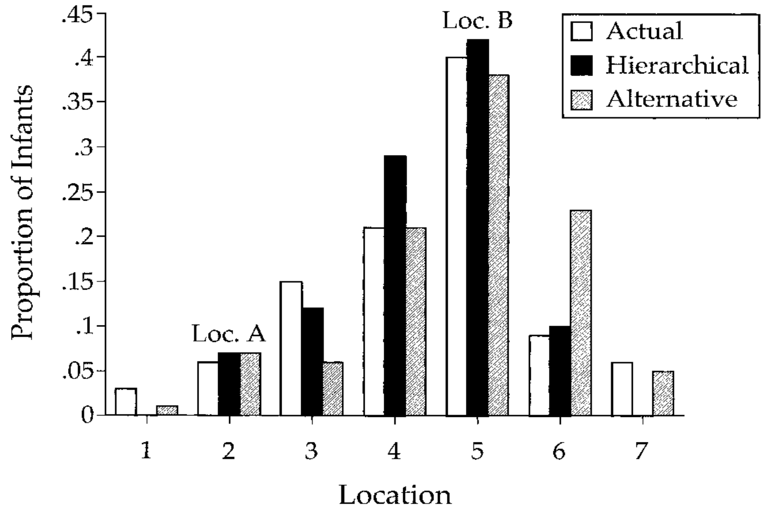

Figure 2 Proportion of infants who searched at each location on Diamond et al.'s (1994) multilocation search task (Actual), the predicted value from the hierarchical competing-systems model (Hierarchical, α = 2 and σ = .6), and the predicted value from the alternative competing-systems model (Alternative, σ_A = 1, σ_B = 1, and σ_I = 1.2).

The hierarchically organized competing-systems account described earlier is similar to the competing-systems account proposed by Diamond et al. (1994). However, Diamond et al. posit that along with memory limitations, poor inhibitory control permits perseverative responding. In the hierarchical competing-systems account, there is no separate inhibition mechanism. Rather, the combined influence of the representational and response-based systems suffices to determine search behavior (cf. Goldman-Rakic, 1987; Pennington, 1994).

A new quantitative model. The hierarchical competing-systems account can be quantified as the sum of a normal distribution and an exponential distribution. The representational system directs an infant's behavior to the B location. The maximum likelihood of the infant's search is at the B location; surrounding locations have a lower probability. This can be expressed as a normal distribution[4] around location B:

$$f(locX) = [1 / \sigma(2\pi)^{1/2}] \exp\{-1/2[(locX - locB) / \sigma]^2\} \qquad (13)$$

Location X is the location at which one wants to calculate the likelihood of search. The parameter in the distribution is σ, whose value depends on the infant's age and the length of the delay. As the infant ages, σ decreases, thereby increasing the probability of searching at the correct location. However, as the length of the delay increases, so does σ, thereby increasing the probability of incorrect search.

After witnessing the first B trial, the infant stores a representation of the hidden object at location B, and the infant will remain oriented to location B. If the influence of the response-based system is strong enough, however, the infant will be pulled toward the A location. Intermediate locations will also elicit search because of their similarity to the B location (cf. Thelen & Smith, 1994). Thus, if the infant is attracted toward location A by the response-based system, it is most likely that the infant will search at the first location away from B and toward A. The next most likely location of search would be the second location away from B and toward A, and so on, until location A is reached. This can be quantified as an exponential distribution.[5]

$$g(locX) = (1/ \alpha) \exp[- |locB — locX| / \alpha], \quad \text{if } locX = locA, \text{ or is between } locA \text{ and } locB \qquad (14)$$

$$g(locX) = 0, \text{ elsewhere, including the situation where } locX = locB. \qquad (15)$$

Again, location X is the location at which one wants to calculate the likelihood of search. The parameter in the distribution is α, whose value may be related to the number of trials at location A. It is important to note that the contribution of the exponential function to the likelihood of search at any location X is at a maximum when α is equal to $|locB - locX|$. For example, if locA = 2, locB = 5, and locX = 3, then the likelihood of searching at location #3 is at a maximum when $\alpha = 2$. We can postulate that $\alpha = K(1 - m / N)$, where m is a constant, and N is the number of A trials. This function has an asymptote at $K = |locB - locA|$.

To implement the model, one must assign values to the parameters σ and α. Then for each location X, one must find h(locX), which is the sum of the normal and exponential functions, namely h(locX) = f(locX) + g(locX). The probability of searching at location X on the first B trial is: $h(locX) / \Sigma_i h(locX_i)$, where i ranges from 1 to the total number of locations.

Diamond et al.'s (1994) data provide a test to see if the proposed distribution can be fitted to data. Recall that in their experiment there were seven hiding locations, that the A location was location #2, and that the B location was location #5. Thus locA = 2 and locB = 5. The values of the parameters are $\sigma = .6$, and $\alpha = 2$ (α_{max} would be 3). As can be seen in Figure 2, which shows the predicted likelihood of search at each location, the hierarchical competing-systems model fits the actual data well.

A novel prediction that can be derived from this model is that as the number of A trials increases, infants are more likely to search at A and less likely to search at intermediate locations. Recall that the likelihood of searching at location X is at a maximum when $\alpha = |locB — locX|$. So, the likelihood of searching at location A is maximized when $\alpha = |locB — locA| = \alpha_{max}$. Thus, with an increase in the number of A trials, a increases toward α_{max} and search is more likely to occur at location A. Consequently, search is less likely to occur at the other locations. To illustrate this prediction, imagine the scenario where locA = 2 and locB = 5. If $\alpha = 1$ (corresponding to few A trials), the likelihood of search due to the response-based system is the highest at location #4, because $\alpha = 1 = |locB - locX|$. As the number of A trials increases, α increases toward $\alpha_{max} = 3$, and search at location A becomes more likely. After a sufficient number of A trials, $\alpha = 3 = \alpha_{max} = |locA - locB|$, and search at location A will be maximized while search at the other incorrect locations will be minimized.

Comparison with an alternative competing-systems model. It may be useful to compare the hierarchical competing-systems model with another quantitative competing-systems model based on Diamond et al.'s (1994) memory plus inhibition theory. It should be noted, however, that Diamond and colleagues do not provide a quantitative model and that alternative quantifications of their theory are possible. The three processes to be considered in our quantification of the memory plus inhibition theory were (1) the tendency to search at location B, (2) the tendency to search at location A, and (3) the inhibition process.

The tendency to search at location B is a direct result of the infant's memory of the object hidden at that location. Following the same logic as in the hierarchical competing-systems model, it seems reasonable to assume that this tendency can be expressed as a normal distribution around location B (see Equation 13). As in the hierarchical competing-systems model, the parameter σ_B may depend on the infant's age and the length of the delay.

The inclination to repeat previously rewarded responses creates a prepotent tendency for the infant to search at location A. It seems reasonable that based on this tendency, the maximum likelihood of search is at the A location, while surrounding locations have lower probabilities. Thus once again, a normal distribution centered around location A seems to be an appropriate quantification. The parameter σ_A reflects the number of trials to the A location.

Diamond et al. (1994) suggest that perseverative responding is due in part to a failure to inhibit the tendency to respond at location A. Therefore, it is necessary to have a quantification of this inhibitory process. Successful inhibition serves to negate the tendency to search at location A, and this process can be modelled as a negatively valued normal distribution centered around location A. The parameter σ_I is a function of the infant's age.

The probability of searching at any location X is based on the sum of these three normal distributions. Younger infants will have very little inhibitory control (i.e., a lower negative-valued inhibitory distribution), which allows for a greater influence of the pre-potent tendency to search at the A location. These younger infants will therefore be likely to search around location A. On the other hand, older infants will have more inhibitory control (i.e., a higher negative-valued inhibitory distribution), which would effectively lower the influence of the infant's tendency to search at location A. In this case, search would more likely occur around location B. When this model was fitted to Diamond et al.'s (1994) data presented earlier, the parameters that led to the best fit were $\sigma_B = 1$, $\sigma_A = 1$, and $\sigma_I = 1.2$. The likelihood of search at each location is shown in Figure 2. One can measure a model's goodness of fit by summing the differences between predicted and observed data. On this measure, the alternative model yields a sum of .29. By comparison, the hierarchical competing-system model yields a sum of .24.

Inspection of Figure 2 allows one to examine further the nature of the differences between these two models. The alternative model appears to violate two of the trends suggested by the data. First, the alternative model over-estimates the proportion of infants who search at location 6 (23% instead of the observed 9%). Second, the alternative model predicts a higher proportion of search activity at location 2 (the A location) than location 3. On the other hand, the hierarchical competing-systems model more accurately predicts the proportion of infants who search at location 6 (10%), and correctly predicts a higher proportion of search activity at location 3 than at the A location. However, the alternative model does succeed in predicting a small amount of search activity at locations 1 and 7, whereas the competing-systems model predicts none. Nevertheless, it appears that overall, the hierarchical competing-system model provides a better account of the observed data. In addition, because the hierarchical competing-systems model fits two parameters whereas the alternative model fits three, it is more parsimonious.[6]

Conclusions

Search paradigms that measure the A-not-B error have been used by researchers to assess different aspects of infants' cognitive development. However, large procedural differences have made it difficult to generalize findings across studies. Meta-analytic techniques allow us to measure simultaneously the effect of a variety of manipulations on search performance. Although meta-analyses have been criticized, most recently by LeLorier, Gregoire, Benhaddad, Lapierre, and Derderian (1997), results from meta-analyses may reveal trends that are counterintuitive or difficult to assess under ordinary conditions. Common criticisms of meta-analyses include the claims that meta-analyses only include published studies, and that they lack strict inclusion criteria. Our analysis, however, did include unpublished studies and our inclusion criteria were strictly defined.

The meta-analysis conducted by Wellman et al. (1986) revealed that while age, delay, and number of locations were significant predictors of infants' search behavior, the number of A trials was not. These results have since influenced the developmental literature. On the one hand, researchers have discussed Wellman et al.'s ". . . counterintuitive finding that infants have often been found to perform better on the AB task when multiple hiding locations are used than when only two hiding locations are used" (Diamond et al., 1994, p. 192). On the other hand, they have reiterated the conclusion that "a single A trial is just as likely to elicit perseveration as several A trials" (Harris, 1989, p. 119; see also Flavell, Miller, & Miller, 1993, p. 63; Hofstadter & Reznick, 1996, p. 648). Clearly, future accounts of the A-not-B error will need to reconsider Wellman et al.'s conclusions in light of the current meta-analysis.

In this article, we presented the results from an updated logistic meta-analysis of the A-not-B error. We concluded that age and the distance between locations were positive predictors of the proportion of infants who searched correctly, while the number of A trials and delay between hiding and search were negative predictors. However, the number of hiding locations negatively predicted the proportion of infants who searched perseveratively, but did not predict the proportion of infants who searched correctly. The two novel findings were: (1) The number of A trials is a significant predictor of infants' search behavior, and (2) the number of hiding locations predicts perseverative but not correct search behavior.

On the basis of these results, we presented a quantitative version of the hierarchical competing-systems model that captures the contribution made to search behavior by two hierarchically arranged dissociable systems. The conscious representational system guides search around the B location (see Equation 13) while the response-based system guides search toward the A location (see Equation 14). Further research is necessary to define more fully the parameterization of the model. For example, we speculate that v is related proportionally to the length of delay, but related inversely to the infant's age. However, the function that predicts v remains to be specified. Nonetheless, the model does make a novel, testable prediction: As the number of A trials increases, the infant is more likely to search at A as opposed to intermediate locations. Future studies are needed to examine the specification of the parameters and the validity of the prediction.

Notes:

1 Hidden units would combine information from two or more input units and the result would be added to the total input strength. Effectively, they would measure interactions. Networks with hidden units were simulated, but because all interactions were nonsignificant, they have been excluded from this article for the sake of simplicity.

[2] The criterion of two sample standard deviations was chosen based on other analyses not mentioned in this article. With this criterion, inclusion as a significant parameter is less likely than if the criterion were a 99% confidence interval.

[3] Theoretically, it is possible that this pattern of results implies that the variable predicts correct behavior, but not perseverative behavior. Separate analyses (not reported here) that used proportion of infants who perseverate as the dependent variable with no covariate confirmed that indeed these variables predicted both correct and perseverative behavior.

[4] According to our criterion, any unimodal symmetrical function could be used; the normal distribution is simply the most common.

[5] According to our criterion, any monotonic increasing function could be used. However, the exponential function has the advantages of being nonlinear and fitting the data well.

[6]It could be argued that the alternative model would produce better results if, following the logic of the hierarchical competing-systems model, an exponential function were fitted for the response-based system instead of a normal distribution. In this situation, the best fit to the data occurred when σ_I was minimal, but the fit was never as good as the more parsimonious hierarchical competing-systems model. Other quantifications of Diamond et al.'s (1994) theory might be more successful, but this remains to be tested.

ACKNOWLEDGMENTS

This research was supported in part by a research grant from NSERC of Canada to Philip David Zelazo. We thank David Bjorklund, Charles Brainerd, Sophie Jacques, Bob Lockhart, and an anonymous reviewer for providing constructive comments on an earlier draft of this manuscript. Special thanks to Geoff Hinton who introduced us to logistic regression with connectionist models.

REFERENCES

References marked with an asterisk indicate studies included in the current meta-analysis.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 38*, 35–44.

*Appel, K. J., & Gratch, G. (1984). Will infants search when "no toy" is hidden? A study of implicit assumptions about the development of object permanence. *British Journal of Developmental Psychology, 2*, 179–188.

Baillargeon, R., Graber, M., Devos, J., & Black, J. (1990). Why do young infants fail to search for hidden objects? *Cognition, 36*, 255–284.

*Bell, M. A., & Adams, S. (1997, April). *Looking vs. reaching: Sex differences in object permanence performance.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
*Bigelow, A. E., MacDonald, D., & MacDonald, L. (1995). The development of infants' search for their mothers, unfamiliar people, and objects. *Merrill-Palmer Quarterly, 41*, 191–208.
*Bjork, E. L., & Cummings, E. M. (1984). Infant search errors: Stage of concept development or stage of memory development. *Memory and Cognition, 12*, 1–19.
*Bremner, J. G. (1978). Spatial errors made by infants: Inadequate spatial cues or evidence of egocentrism? *British Journal of Psychology, 69*, 77–84.
*Bremner, J. G., & Bryant, P. E. (1977). Place versus response as the basis of spatial errors made by young infants. *Journal of Experimental Child Psychology*, *23*, 162–171.
*Butterworth, G. E. (1977). Object disappearance and error in Piaget's stage IV task. *Journal of Experimental Child Psychology*, *23*, 391–401.
*Butterworth, G. E., Jarrett, N., & Hicks, L. (1982). Spatiotemporal identity in infancy: Perceptual competence or conceptual deficit? *Developmental Psychology*, *18*, 435– 449.
*Cummings, E. M., & Bjork, E. L. (1983a). Perseveration and search on a five-choice visible displacement hiding task. *Journal of Genetic Psychology*, *142*, 283–291.
*Cummings, E. M., & Bjork, E. L. (1983b). Search behavior on multi-choice hiding tasks: Evidence for an objective conception of space in infancy. *International Journal of Behavioral Development, 6*, 71–87.
Dempster, F. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*, 45–75.
Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development, 56*, 868–883.
Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and knowledge* (pp. 67–110). Hillsdale, NJ: Erlbaum.
*Diamond, A., Cruttenden, L., & Neiderman, D. (1994). AB with multiple wells: 1. Why are multiple wells sometimes easier than two wells? 2. Memory or memory + inhibition? *Developmental Psychology*, *30*, 192–205.
*Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, *74*, 24–40.
*Evans, W. F. (1973). *The stage IV error in Piaget's theory of object concept development: An investigation of the role of activity*. Unpublished dissertation proposal, University of Houston, Houston, TX.
*Evans, W. F., & Gratch, G. (1972). The stage IV error in Piaget's theory of object concept development: Difficulties in object conceptualization or spatial location? *Child Development*, *43*, 682–688.
Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive development* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
Frye, D. (1980). Stages of development: The stage IV error. *Infant Behavior and Development*, *3*, 115–126.
Goldman-Rakic, P. S. (1987). Cortical circuitry and cognitive function. *Child Development, 58*, 601–622.
*Gratch, G., Appel, K. J., Evans, W. F., LeCompte, G. K., & Wright, N. A. (1974). Piaget's stage IV object concept error: Evidence of forgetting or object conception. *Child Development*, *45*, 71–77.
Harnishfeger, K. K., & Bjorklund, D. F. (1993). The ontogeny of inhibition mechanisms: A renewed approach to cognitive development. In M. L. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development* (pp. 28–44). New York: Springer-Verlag.
*Harris, P. L. (1973). Perseverative errors in search by young infants. *Child Development*, *44*, 28–33.
Harris, P. L. (1974). Perseverative search at a visibly empty place by young infants. *Journal of Experimental Child Psychology*, *18*, 535–542.
Harris, P. L. (1983). Infant cognition. In M. M. Haith & J. J. Campos (Eds.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 689–782). New York: Wiley.
Harris, P. L. (1986). Bringing order to the A-not-B error. *Monographs of the Society for Research in Child Development*, *51*(3, Serial No. 214), 52–61.

Harris, P. L. (1989). Object permanence in infancy. In A. Slater & G. Bremner (Eds.), *Infant Development* (pp. 103– 121). Hillsdale, NJ: Erlbaum.
*Hofstadter, M., & Reznick, J. S. (1996). Response modality affects human infant delayed-response performance. *Child Development, 67*, 646–658.
*Horobin, K. M., & Acredolo, L. P. (1986). The role of attentiveness, mobility history, and separation of hiding sites on stage IV search behavior. *Journal of Experimental Child Psychology, 41*, 114–127.
Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century.
Hull, C. L., Felsinger, J., Gladstone, I., & Yamaguchi, H. (1947). A proposed quantification of habit strength. *Psychological Review, 54*, 237–254.
Hunter, W. S. (1913). The delayed reaction in animals and children. *Behavior Monographs, 2*, 1–86.
Hunter, W. S. (1917). The delayed reaction in a child. *Psychological Review, 24*, 74–87.
*Landers, W. F. (1971). Effects of differential experience on infants' performance in a Piagetian Stage IV object-concept task. *Developmental Psychology, 5*, 48–54.
LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized control trials. *New England Journal of Medicine, 337*, 536–542.
Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior.* J. Tizard (Ed.). New York: Pergamon Press.
Marcovitch, S. (1996). *The effect of number of pre-switch trials on 2-year-olds' perseveration in a multi-step multi-location search task.* Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.
*Mistry, R., Colombo, J., & Saxon, T. (1997, April). *Stage-IV "A-not-B error" and the facilitation of auditory cues.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
Munakata, Y. (1997). Perseverative reaching in infancy: The roles of hidden toys and motor history in the AB task. *Infant Behavior and Development, 20*, 405–416.
Neter, J., & Wasserman, W. (1974). *Applied linear statistical models.* Homewood, IL: Irwin.
Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.
Pennington, B. F. (1994). The working memory function of the prefrontal cortices: Implications for developmental and individual differences in cognition. In M. Haith, J. B. Benson, R. J. Roberts, Jr., & B. F. Pennington (Eds.), *The development of future oriented processes* (pp. 243–289). Chicago: University of Chicago Press.
Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.
*Roberts, J. (1997, April). *Attention and mood are related to 8- and 10-month-olds' performance on the A-not-B task.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
*Schuberth, R. E., Werner, J. S., & Lipsitt, L. P. (1978). The stage IV error in Piaget's theory of object concept development: A reconsideration of the spatial localization hypothesis. *Child Development, 49*, 744–748.
*Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review, 106*, 235–260.
*Sophian, C. (1985). Perseveration and infant's search: A comparison of 2- and 3-location tasks. *Developmental Psychology, 21*, 187–194.
*Sophian, C., & Sage, S. (1985). Infants' search for hidden objects: Developing skills for using information selectively. *Infant Behavior and Development, 8*, 1–14.
*Sophian, C., & Wellman, H. M. (1983). Selective information use and perseveration in the search behavior of infants and young children. *Journal of Experimental Child Psychology, 35*, 369–390.
*Sophian, C., & Yengo, L. (1985). Infants' search for visible objects: Implications for the interpretation of early search errors. *Journal of Experimental Child Psychology, 40*, 260–278.
Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Thorndike, E. L. (1921). *Educational psychology: Briefer course.* New York: Teachers College Press.
van Camp, D. (1996). *A user's guide for the Xerion neural network simulator (Version 4.1).* Toronto: University of Toronto, Department of Computer Science.
*Webb, R. A., Massar, B., & Nadolny, T. (1972). Information and strategy in the young child's search for hidden objects. *Child Development, 43*, 91–104.
Wellman, H. M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. *Monographs of the Society for Research in Child Development, 51*, 1–51, 62–67.
Williams, S. B. (1938). Resistance to extinction as a function of the number of reinforcements. *Journal of Experimental Psychology, 23*, 506–521.
*Wishart, J. G. (1987). Performance of young nonretarded children and children with Down Syndrome on Piagetian infant search tasks. *American Journal of Mental Deficiency, 92*, 169–177.
Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.), *Language structure, discourse and the access to consciousness* (pp. 113–153). Amsterdam: John Benjamins.
Zelazo, P. D., Reznick J. S., & Spinazzola, J. (1998). Representational flexibility and response control in a multistep multilocation search task. *Developmental Psychology, 34*, 203–214.
Zelazo, P. R., & Zelazo, P. D. (1998). The emergence of consciousness. In H. H. Jasper, L. Descarries, V. F. Castellucci, & S. Rossignol (Eds.), *Consciousness: At the frontiers of neuroscience, advances in neurology* (Vol. 77, pp. 149– 165). Philadelphia: Lippincott-Raven.